E

The International Treaty
ON PLANT GENETIC RESOURCES FOR FOOD AND AGRICULTURE

INTERNATIONAL TREATY ON PLANT GENETIC RESOURCES FOR FOOD AND AGRICULTURE

FIRST MEETING OF THE EXPERT CONSULTATION ON THE GLOBAL INFORMATION SYSTEM ON PLANT GENETIC RESOURCES FOR FOOD AND AGRICULTURE

San Diego, USA, 7-8 January 2015

Compilation of Submissions Received from Contracting Parties, Non-Contrating Parties and International Relevant Organizations

---

Two contributions to an integrated, global, accession-level information system for ex situ conservation

---

The inputs provided in this addendum were received on 2 January 2015 by the Secretary of the Governing Body of the International Treaty on Plant Genetic Resources for Food and Agriculture and they are presented in the language and form in which they were receveid.

Two contributions to an integrated, global, accession-level information system for *ex situ* conservation

Input Paper to the ITPGRFA Consultation on the Global Information System on Plant Genetic Resources for Food and Agriculture (COGIS-PGRFA)

Provided by: The Global Crop Diversity Trust

Introduction

Approximately 7 million PGRFA accessions are currently conserved in genebanks worldwide. The data associated with these accessions is critical to both the enhanced use of the collections and the development of an effective, rational global *ex situ* conservation system. Article 17 of the International Treaty on Plant Genetic Resources for Food and Agriculture calls for the development and strengthening of a global information system for Plant Genetic Resources for Food and Agriculture (PGRFA), based on existing information systems. Genesys and DivSeek are two contributions to the management and sharing of accession-level information, on which the Crop Trust is collaborating with the Secretariat of the International Treaty.

Genesys (https://www.genesys-pgr.org) is a global web portal bringing together different genebank databases with the aim of eventually providing a one-stop shop for potential users of conserved germplasm. Genesys focuses on passport data, though it also has provision for handling basic characterization and evaluation data. However, recent advances in genomics and bioinformatics hold the promise to truly unleash the value of genebanks. The recently launched Diversity Seek initiative (DivSeek http://www.divseek.org/) will develop an information management platform to provide ready access to both genotypic and phenotypic data associated with genebank accessions.

Genesys

The development of Genesys was initiated by the Global Crop Diversity Trust in collaboration with the Secretariat of the International Treaty in 2006, with funding mainly from the Bill & Melinda Gates Foundation. Bioversity International implemented and released the first version of the portal in 2009. This contained the 2.33 million accession records that were available in the SINGER, Eurisco and GRIN information systems at the time. Genesys is the successor to, and an expansion of, SINGER, the collective information system of the CGIAR genebanks. The SINGER website was phased out in 2013, and now Genesys is the collective way the CGIAR genebanks share their passport data. A second version of the portal was launched by the Crop Trust in 2014. Genesys now contains 2.77 million accession records and features seamless data integration with a number of other information systems. A few more genebanks have become data providers, and the data of the original providers has been updated.

A number of lessons have been learned during Genesys development that could be of value to discussion of GIS-PGRFA. First, data providers needed to feel secure that once their data was published in Genesys, it would be safe from adulteration and there would be proper attribution when it was used. Even though all the data now in Genesys was available online in public databases, there was still a need to develop a formal data sharing agreement.

Another important lesson was the need to consider carefully the governance of the system. While the Crop Trust physically hosts Genesys and coordinates its development, a group of international experts, including a representative of the International Treaty Secretariat, has been constituted to provide technical advice and strategic guidance (the Oversight Committee). Programming is outsourced to private companies by competitive calls or direct commission. All source code is licensed under Apache License 2. The source code repositories are accessible at https://bitbucket.org/genesys2/. Feedback about the interface and functionality is being elicited from users. Mechanisms are also being considered to fully involve interested data providers in the further development of the portal.

Another lesson has been not to re-invent the wheel. Genesys integrates with, or uses data from, the following other information systems:

- FAO World Information and Early Warning System[1] (WIEWS)
- GeoNames[2] geographical database
- Svalbard Global Seed Vault database[3]
- MapQuest[4] for world maps
- ITPGRFA's Easy-SMTA

Flexibility has been critical. The original manual method of upload to Genesys has been automated, but this has required a variety of different tools and approaches. It has also required a much expanded help desk function. For many genebanks, the lack of an adequate data management systems is a bottleneck for participation in Genesys. The adoption of software such as GRIN-Global will address this, and another key lesson is thus that expanding the base of data providers requires investment. There can be no development of global information systems without the development of local genebank information systems. Publishing data on Genesys is the ultimate step in the development of a genebank information system, showing the world that the genebank data adheres to agreed international standards (most notably MCPD[5]). Through Genesys, genebanks are able to share experiences, develop showcases, and share ideas on enhancing data quality, leading to a virtuous circle of improvement.

In addition to accession-level data, Genesys currently also provides the following views and functionalities, of interest to different types of users, ranging from genebank managers to researchers to policy makers:

- Institute information and summary overviews of holdings
- Country level summary
- Crop specific access to accession lists
- Drill-down data filtering
- Integrated map tile server
- Ordering system

---

[1] http://apps3.fao.org/wiews

[2] http://www.geonames.org/

[3] http://nordgen.org/sgsv/

[4] http://developer.mapquest.com/

[5] MCPD, http://www.bioversityinternational.org/e-library/publications/detail/faobioversity-multi-crop-passport-descriptors-v2-mcpd-v2/

In the current ecosystem of PGRFA information, Genesys is positioned as the source of the next level of genebank data after WIEWS. WIEWS serves as the authority for genebank identification codes used in data exchange and provides statistics on the numbers of accessions of different species in all the world's genebanks. Genesys currently hosts accession-level data of an estimated 25% of the world's *ex situ* accessions, with expansion to other genebanks the first priority. The Secretariat of ITPGRFA should consider leveraging policy and operational support to Genesys in order to reach a wider set of data providers, users and partners, and be recognised as one component of the Global Information System, alongside WIEWS. Together, these could potentially form an important pillar of GIS-PGRFA, to which other modules could be aggregated. One such is DivSeek.

DivSeek

The Diversity Seek (DivSeek) Initiative has the mission to enable genebanks, breeders and researchers to better manage and mobilize the vast range of plant genetic variation to accelerate the rate of crop improvement. Imagine that all accession of a given crop in genebanks around the world were fully sequenced, and that the massive amount of resulting data was readily available to genebank managers and breeders through their browsers. This would mean that complex management tasks such as the identification of duplicates and near-duplicates within and among genebanks would suddenly be not just feasible but simple. Core collections could be produced with a click. And, given phenotypic data on a well-chosen subset of accessions, predictions could be made about the performance of accessions that had not yet been expensively evaluated. DivSeek seeks to make this vision possible.

As a first step, DivSeek will agree a system of standards, protocols, tools, resources and best practices for generating, organizing, structuring, indexing, retrieving, sharing and analyzing linked genotypic and phenotypic data on accessions. Next, DivSeek will build an information platform providing access to genotypic, phenotypic, and other types of information linked to germplasm. This will be complemented by a suite of computational tools and resources that would allow diverse users to visualize and analyze the data for multiple purposes.

Developing such an information platform is a dauntingly ambitious task, and will require an approach that is sufficiently flexible to allow it to adapt to new requirements and technologies, whilst maintaining a set of core standards that allow data exchange, integration and interoperability. To achieve this, a substantial shift in mindset is needed. Genebank managers, crop scientists and plant breeders need to feel comfortable handling huge quantities of diverse, high-dimensional data. DivSeek will therefore promote interdisciplinary capacity building programs to help form the next generation of bioinformatics-savvy plant scientists.

DivSeek will provide an 'umbrella' under which different partners can work together towards common goals. There are already well-established players with their germplasm, genomic and phenotypic data, and governance structures. DivSeek will aim to bring together these existing resources while developing new structures for access, information sharing and innovation. A landscaping analysis will be done initially to record the breadth and depth of relevant efforts in the field, identify potential partners and begin a dialogue with the community. The landscaping will document not only the technical aspects of existing systems, but also their governance structures and policies with regard to data access and use.

To initiate the creation of a consortium-like structure, interested potential partners

are being asked to sign an expression of interest letter in which they agree to commit in principle to working together in DivSeek. Thirty-two have so far done so. Once partners have been identified and are on board, DivSeek will focus on a set of pilot projects which will assess the type of data that the platform will need to manage; describe how data are currently used, accessed, analyzed and integrated; and develop solutions for improved and increased use of the data.

Based on the experience gained in the pilot projects, a global platform will be developed providing transparent, integrated access to genomic and phenomic data linked to MLS germplasm, and accompanied by a suite of tools and resources. The platform will comprise data standards, structures and sharing protocols that will govern the way sequencing data (generated using diverse platforms), phenotyping data (organismal, biochemical, molecular etc.) and environmental data (field, greenhouse, growth chamber etc.) are annotated, accessed, shared and integrated by all users. Mechanisms for governance, communications and IP management will be assessed and agreed.

An interim facilitation unit has been established while a permanent governance structure is developed by the DivSeek community. The Global Crop Diversity Trust hosts and implements the interim facilitation unit jointly with the Secretariat of the International Treaty, and operates it with additional inputs by the CGIAR Consortium, the Global Plant Council and other experts/organizations. Hosting arrangements for the facilitation unit will be formalized once the DivSeek consortium is officially established. The secretariat function of the consortium may continue to be hosted with the same modalities.

Summary

The experience accumulated so far by the Crop Trust with Genesys, and the concepts underpinning DivSeek, highlight the following issues that it might be advantageous for the Global Information System to take into consideration from the earliest stages of development:

- agree a robust yet flexible collaboration and governance mechanism with all stakeholders
- inventory and build on existing standards and structures
- ensure scalability
- address capacity building